Methodology Minor Field Exam

Fall 2019

For the minor field exam, you must answer two questions, one in the morning session and one in the afternoon session. In the afternoon session, you may use the software of your choice. You are free to use whatever word processing or typesetting software you like to write your answers. The questions must be answered in the allotted time.

For the morning session, internet usage is prohibited. For the afternoon session, you may use the internet to download software packages or look up reference information as you complete the data analysis. Your work must be fully your own. Enjoy this opportunity to showcase your skills.

Morning Session: Statistical Theory and Modeling Decisions

Answer one of the following three questions:

1. *Experimental Methods:* To what extent are the following experimental design choices a problem for the estimation of the average treatment effect (ATE)?

   In your answer, define the ATE and specify how each design choice affects the researcher?s analyses as well as the interpretation of the results. Illustrate each case with (hypothetical or real) examples from political science research.

   A : The researcher has two treatment groups but no control group.

   B : The researcher cannot fully control which subjects are exposed to the treatment and which are not.

   C : The researcher randomly assigns subjects to experimental and control groups, but has post-test (after the intervention) measurements only, no pre-test measurements.

   D : The researcher cannot do random assignment of subjects to treatment and control groups and chooses to do matching instead.

   E : The researcher uses a convenience sample for the experiment.

2. *Bayesian Statistics:* An essential component of Bayesian inference is that the research must formally specify priors for all model parameters. While some scholars using frequentist or likelihoodist methods may object to the insertion of researcher judgment at this stage, many Bayesians would argue that priors are a strength of the Bayesian approach. What are some of the main objections to the use of priors? How can priors can be beneficial for inference? Give one example of when one might use a prior for reasons other than inserting researcher judgment.

   Describe common approaches to formulating priors for the parameters of a Bayesian model. Explain the process of constructing the prior. What are the relative strengths and weaknesses of the different approaches?

   Describe a real or imagined dataset that you might want to model using Bayesian statistics. Your example data can have any kind of features you think would be illustrative (e.g., pure cross-section, time series, panel data, geospatial data, multilevel data, etc.). Suppose you were specifying priors for your example data: Describe what you would do specifically if you followed one of the common approaches described earlier. Why is the approach you are choosing the best option?

3. *Linear Model Theory:* When estimating a linear model via OLS, one assumes the Gauss-Markov assumptions to be true in order for the estimates to be BLUE. What are these Gauss-Markov assumptions? What does it mean for an estimator to be BLUE? What are the most common violations of Gauss-Markov assumptions and what types of data are most likely to lead to violations of these assumptions? What are the implications to violating Gauss-Markov, specifically in terms of interpreting model results? Finally, which assumption do you feel is the most important to not violate and why?

Afternoon Session: Analyzing Data

Answer one of the following two questions:

4. *Linear Regression:* Please analyze the dataset *engagement.dta* using a linear regression model. The dataset contains information on overall civic engagement during the 2012 presidential election. The variable of interest, **engscale**, is a scale measuring respondents' civic engagement (dependent variable; 0 = doesn't participate in any activity, 5 = participates in all activities).

   The input variables are as follows (you must use them all):

   education : Educational attainment (1 = didn't complete high school, 6 = post-graduate degree).

   income : Household income (1 = less than $10,000, 18$=250,000 or more).

   age : Respondent's age.

   stghpid : A measure of strength of partisanship (0 = Independent, 3 = strong party ID).

   ideodist : Ideological distance to President Obama (0 = same position as President Obama, 6 = opposite extreme relative to President Obama).

   Present the results of this model in a table including the coefficients, the standard errors, the R^2, and any additional information you would like. What can you conclude from the t-ratios associated with each coefficient? What can you conclude from the model fit?

   Please test the conditional hypothesis that educational attainment fosters civic engagement, and that the positive influence of education is especially intense for individuals with strong partisan attachments. Estimate a new model to test this hypothesis and discuss the results. Illustrate the nature of this conditioned relationship by graphing predicted values and confidence intervals. Provide a detailed interpretation of the conditional relationship and whether or not you think it matters. Compare the fit of the two models and discuss the implications of including the conditional relationship described above relative to not including this. Which model do you feel is a better fit to the data and why?

   Assess whether or not there are problems with collinearity and heteroskedasticity. Include the appropriate graphs or tables and be sure to discuss the results of these tests in detail. Finally, discuss whether or not you think OLS is the appropriate estimator for these data. If so, justify your response. If not, what model do you think would be a better estimator and why?

5. *Poisson Regression:* Please analyze the dataset *couart2.dta* using a poisson regression model. The dataset contains information on the number of articles published by PhD students during the last 3 years of their education. The variable of interest, **art**, is the number of articles published by students in the last 3 years they attend a PhD program.

   The input variables (you must use them all) are:

   fem : Dummy for gender (1 = female).

   mar : Dummy for marital status (1=married).

   kid5 : Number of children

   phd : Prestige score of PhD granting institution (higher = more prestigious).

   ment : Number of articles published by student's mentor in last 3 years.

   Present the results of this model in a table including the coefficients, the standard errors, and any additional information you would like. What can you conclude from the t-ratios associated with each coefficient? What can you conclude from the model fit?

   Present graphs of predicted counts against all covariates.

   Finally, discuss whether or not you think Poisson regression is the appropriate technique for these data and justify your answer. If no, then discuss other options and why they may be more appropriate.